

## REMARKS

Prior to this response, Claims 1-26 were pending in the application. By this response, Claims 1, 2, 7, 13-15, 20 and 26 are amended. No claims are added or canceled. Hence, Claims 1-26 are presently pending in the application.

## SUMMARY OF THE REJECTIONS

Claims 1-13 were rejected under 35 U.S.C. §101 as allegedly directed to non-statutory subject matter.

Claims 1, 13, 14 and 26 were rejected under 35 U.S.C. §102(b) as allegedly anticipated by Gal et al. (“*Gal*”; U.S. Pat. No. 5,729,732). Claims 2-4, 7, 15-17 and 20 were rejected under 35 U.S.C. §103(a) as allegedly unpatentable over *Gal* in view of Poosala (“*Poosala*”; “Improved Histograms for Selectivity of Range Predicates”). Claims 5 and 18 were rejected under 35 U.S.C. §103(a) as allegedly unpatentable over *Gal* in view of Ogi (“*Ogi*”; U.S. Pat. No. 5,854,938). Claims 6 and 19 were rejected 35 U.S.C. §103(a) as allegedly unpatentable over *Gal* and *Poosala*, further in view of Marks (“*Marks*”; U.S. Pat. No. 5,748,844).

Claims 8-12 and 21-25 were objected to as being dependent upon a rejected base claim, with an indication of allowability if rewritten in independent form.

## REJECTIONS NOT BASED ON THE PRIOR ART

### Rejection under 35 U.S.C. §101

Pursuant to the suggestion provided in the Office Action, independent Claims 1, 2 and 7 are amended herein to recite a “[a] computer-implemented method ...” rather than “[a] method ...”. Therefore, it is clear that the steps recited in each of these claims are performed by a computer, such as “computer system 100” of FIG. 1, rather than mentally in conjunction with pen and paper. Claims 3-6 and 8-13 depend directly or indirectly from one of Claims 1, 2 and

7. These claims inherit the limitations of the claims from which they depend and, therefore, are also computer-implemented method claims. The rejection of Claims 1-13 based on 35 U.S.C. §101 is now moot, and withdrawal of this rejection is respectfully requested.

## REJECTIONS BASED ON THE PRIOR ART

### I. Rejection under 35 U.S.C. §102(b)

Claims 1, 13, 14 and 26 were rejected under 35 U.S.C. §102(b) as allegedly anticipated by *Gal*. This rejection is traversed.

For a proper anticipation rejection, a reference must show each and every feature of a claim in the same combination as claimed. *Connell v. Sears, Roebuck & Co.*, 722 F.2d 1542, 1548, 220 USPQ 193, 198 (Fed. Cir. 1983). The Office Action relied on *Gal* for an alleged showing of the following, with respect to the rejection of Claim 1:

randomly selecting data items from said particular set of data to produce a sampled set of data items;

determining a range for each bucket of said plurality of buckets based on a number of the distribution key values associated with said sampled set of data items that fall within said range.

However, *Gal* does not disclose these features and, therefore, *Gal* does not anticipate Claim 1.

A review of an example described in the present specification may serve as an aid to understanding the embodiment recited in Claim 1. A manner in which bucket ranges can be determined based on the number of distribution key values associated with the sampled set of data items that fall within a range is described in the specification, by way of example, at page 9, line 14 through page 10, line 9. In the example, the goal is to distribute data items from a

particular set of data evenly to four buckets, where the set of data includes data for persons having ages from 0 to 99 (i.e., age is the distribution key), and a sample of 200 data items is randomly selected from the set. Hence, in this example, the ranges to use to distribute the data items to the four buckets are determined for each bucket, based on the sample, in a manner that would equally distribute the data into the buckets. Equally distributing the data among the buckets is a matter of user choice and the actual criteria for selecting the ranges for the buckets could be different from this example. However, the more equal the distribution among the buckets, typically, the more efficient are subsequent parallel operations performed on the data items (see page 1, line 20 through page 2, line 4).

With (a) the goal of equal distribution across buckets, with (b) 4 buckets, and with (c) 200 sample data items, it is clear that the 4 bucket ranges would be determined based on ranges that each encompasses approximately 50 (200 samples/4 buckets) sampled data items. Hence, each range should encompass approximately 25% of the sampled data items, as described at page 9, line 25 through page 10, line 3. The ranges are determined based on what ranges result in each of the 4 buckets encompassing 25% of the sampled data items. Because, in the example, 25% of the sampled data items fall within a range of 0-14; 25% fall within the range of 15-40; 25% fall within the range of 41-70; and 25% fall within the range of 71-99, the ranges to assign to each bucket are determined to be 0-14, 15-40, 41-70 and 71-99.

Differences between the disclosure of *Gal* and what is recited in Claim 1 are as follows. First, *Gal* takes a sample of key values, during a preprocessing phase, in order to estimate the range of key values in a file. This estimated range is then divided into subintervals which serve as buckets in a subsequent sorting phase. Col. 3, lines 22-25. By distinction, *Gal* does **not** disclose that this sample of key values is used to determine a respective range for each of a

plurality of buckets, as in Claim 1. Estimating a single range of key values based on a sample group of records (e.g., a file) is different than determining multiple respective ranges for multiple respective buckets. *Gal* goes on to disclose how the range estimate is calculated based on minimum and maximum key values from the sample, and how the number of buckets (i.e., subintervals) is calculated. Col. 3, line 46 through col. 4, line 39.

Next, *Gal* sorts the items from which the sample was taken by calculating an index of the subinterval to which each item of the file belongs (col. 4, lines 61-65). This seems to mean distributing the items to buckets for sorting purposes, in a divide-and-conquer type of sorting routine (e.g., quicksort) in which items are placed in groups and then sorted within each group. The items are distributed into buckets based on the leading 4 bytes of a byte-representation of the respective key values, so that all key values in each bucket share the same 4-byte prefix which can be ignored for purposes of sorting.

Col. 6, lines 7-15 of *Gal* describes a special case in which more than half the key values in the items to be sorted share the same 4-byte prefix. In such a case, this 4-byte common prefix is effectively used as the pivot in the divide-and-conquer sort routine, by forming groups of items based on whether the first 4 bytes of an item's key value is smaller than, equal to, or larger than the common prefix. The Office Action contends that the foregoing teaching is a teaching of "determining a range for each bucket of said plurality of buckets based on the number of the distribution key values associated with said sampled set of data items that fall within said range" (underline added). *Gal* does not provide such a teaching.

*Gal* determines ranges for three buckets loosely based on the 4-byte common prefix for key value representations, presumably in conjunction with minimum and maximum key values from the set of items to be sorted. Thus, key values for one bucket would presumably range between the minimum value and the value(s) that is represented by the 4-byte common prefix;

key values for another bucket would contain all the key values represented by the 4-byte common prefix; and key values for another bucket would presumably range between the value(s) that is represented by the 4-byte common prefix and the maximum value. The number of values that fall within each range is a simple fall-out from establishment of the ranges based on (1) the 4-byte common prefix, and (2) the corresponding relation of key value byte-representations, for the entire group of items being sorted, to the 4-byte common prefix.

Hence, the only determination that is based on the number of values falling within a given range is the decision to establish one bucket in association with a 4-byte prefix that is common to “more than half” the items in the entire group. Consequently, this *Gal* scheme (1) does **not** determine the ranges (e.g., the range of the “equal to” bucket is the range of values whose first 4 bytes of a corresponding byte-representation are equal to the 4-byte common prefix) based on the number of values that fall within those ranges, rather, the “equal to” bucket range is a mere fall-out resulting from the decision to establish such a bucket. Significantly, neither of the ranges for the “less than” or “greater than” buckets is based on the number of key values falling within those ranges. Rather, either of these two buckets could, in fact, have zero key values falling within the range if there are no key values whose first 4 bytes of a corresponding byte-representation are less than, or are greater than, the 4-byte common prefix.

Furthermore, this *Gal* scheme (2) does **not** determine bucket ranges based on randomly selected sample of items from a set of items being distributed to buckets, rather, it is based on the entire set of items. Still further, this *Gal* scheme (3) is **not** based on distribution key values, per se, but on a byte representation of the value. Based on the foregoing, *Gal* does not teach or fairly suggest that a range is determined for each bucket based on the number of key values, associated with the sampled set, that fall within the corresponding range.

For the foregoing reasons, a *prima facie* case of anticipation is not established because *Gal* does not disclose each and every element of Claim 1. Therefore, withdrawal of the rejection of Claim 1 under 35 U.S.C. §102(b) is kindly requested.

Claim 13 depends from Claim 1 and, therefore, is patentable over *Gal* for at least the same reasons as Claim 1. Claims 14 and 26 are product claims that include features that correspond to Claims 1 and 13, respectively. Thus, Claims 14 and 26 are patentable over *Gal* for at least the same reasons as Claim 1.

II. Rejections under 35 U.S.C. §103(a)

II.A. Claims 2-4, 7, 15-17 and 20

Claims 2-4, 7, 15-17 and 20 were rejected under 35 U.S.C. §103(a) as allegedly unpatentable over *Gal* in view of *Poosala*. This rejection is traversed.

Claim 2 recites similar steps as recited in Claim 1, such as randomly selecting data items ... from a particular set of data and determining ranges based on the number of distribution key values associated with the selected data items, both of which are already discussed above in reference to Claim 1. The discussion of these steps in view of *Gal*, in reference to Claim 1, applies equally to Claim 2 and shows why a *prima facie* case of obviousness is not established with respect to Claim 2. In addition, *Poosala* does not cure these deficiencies in the teachings of *Gal*.

Claims 3 and 4 depend from Claim 2, and Claim 15-17 are product claims that include features that correspond to Claims 2-4, respectively. Thus, Claims 2-4 and 15-17 are patentable over *Gal* in view of *Poosala* for at least the same reasons as Claim 2.

Claim 7 recites, among others, similar steps as recited in Claim 1, such as randomly selecting data items from a particular set of data and determining ranges based on the number of distribution key values associated with the selected data items, both of which are already discussed above in reference to Claim 1. The discussion of these steps in view of *Gal*, in reference to Claim 1, applies equally to Claim 7 and shows why a *prima facie* case of obviousness is not established with respect to Claim 7. Furthermore, *Poosala* does not cure these deficiencies in the teachings of *Gal*.

Claim 7 further recites randomly selecting durable storage units and using the data items stored on the randomly selected storage units as the sampled set of data items. The Office Action relies on Sections 7.1 and 7.2 of *Poosala* for an alleged teaching of this feature. However, *Poosala* in no way teaches or fairly suggests the random selection of data items by randomly selecting storage units and using the data items stored thereon as the basis for a sample of data items. In other words, the basis for the random selection of data items recited in Claim 7 is the random selection of storage units and the data stored thereon. The reservoir sampling scheme used in *Poosala* randomly samples  $n$  tuples for insertion in a “reservoir”, skipping a random number of records, replacing a random tuple in the reservoir with the next tuple, and repeating. This reservoir sampling scheme does not mention any random sampling of items based on a random sampling of storage units on which some of the items are stored. For this additional reason, Claim 7 is patentable over the teachings of *Gal* and *Poosala*.

Claim 20 is a product claim that includes features that correspond to Claim 7. Thus, Claim 20 is patentable over *Gal* in view of *Poosala* for at least the same reasons as Claim 7.

Withdrawal of the rejection of Claims 2-4, 7, 15-17 and 20 under 35 U.S.C. §103(a) is requested.

II.B. Claims 5 and 18

Claims 5 and 18 were rejected under 35 U.S.C. §103(a) as allegedly unpatentable over *Gal* in view of *Ogi*. This rejection is traversed.

Claim 5 depends from Claim 1 and Claim 18 is a product claim that includes features that correspond to Claim 5. The discussion of the deficiencies in the teachings of *Gal*, in reference to Claim 1, applies equally to its dependent claims and shows why a *prima facie* case of obviousness is not established with respect to these claims. Furthermore, *Ogi* does not cure these deficiencies in the teachings of *Gal*. Thus, Claims 5 and 18 are patentable over *Gal* in view of *Ogi* for at least the same reasons as Claim 1. Withdrawal of the rejection of Claims 5 and 18 under 35 U.S.C. §103(a) is requested.

II.C. Claims 6 and 19

Claims 6 and 19 were rejected under 35 U.S.C. §103(a) as allegedly unpatentable over *Gal* and *Poosala* further in view of *Marks*. This rejection is traversed.

Claim 6 depends from Claim 2 and Claim 19 is a product claim that includes features that correspond to Claim 6. The discussion of the deficiencies in the teachings of *Gal*, in reference to Claim 2, applies equally to its dependent claims and shows why a *prima facie* case of obviousness is not established with respect to these claims. Furthermore, *Marks* does not cure these deficiencies in the teachings of *Gal* and *Poosala*. Thus, Claims 6 and 19 are patentable over *Gal* in view of *Poosala* further in view of *Marks* for at least the same reasons as Claim 2. Withdrawal of the rejection of Claims 6 and 19 under 35 U.S.C. §103(a) is requested.

CONCLUSION

For at least the reasons set forth above, it is respectfully submitted that the pending claims (1-26) are in condition for allowance. Therefore, the issuance of a formal Notice of Allowance is believed next in order, and that action is most earnestly solicited.

The Examiner is respectfully requested to contact the undersigned by telephone if it is believed that such contact would further the examination of the present application.

Please charge any shortages or credit any overages to Deposit Account No. 50-1302.

Respectfully submitted,

HICKMAN PALERMO TRUONG & BECKER LLP

Date: 3/9/05

John D. Henkhaus  
John D. Henkhaus  
Reg. No. 42,656

2055 Gateway Place, Suite 550  
San Jose, CA 95110  
(408) 414-1080  
Facsimile: (408) 414-1076

**CERTIFICATE OF MAILING**

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Mail Stop Amendment,  
Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450

on 3/9/05

by

Darci Sakamoto  
Darci Sakamoto